COURSE CODE	COURSE TITLE	L	T	P	C
1152CS139	DATA SCIENCE	3	0	0	3

Course Category: Program Elective

A. Preamble:

This course introduces to the concepts of Data Science.

B. Prerequisite Courses:

Sl. No	Course Code	Course Name
1	1151CS107	Data Base Management System

C. Related Courses:

Sl. No	Course Code	Course Name
1	1156CS601	Minor Project
2	1156CS701	Major Project

D. Course Outcomes:

Upon the successful completion of the course, students will be able to:

CO Nos.	Course Outcomes	Knowledge Level (Based on revised Bloom's Taxonomy)
CO1	Explain the basics of big data analytics life cycle	K2
CO2	Outline the data analysis concepts using R	K2
CO3	Interpret the association rules and regression technique	K2
CO4	Explain the concepts of classification and time series analysis	K2
CO5	Relate the text analysis and Hadoop map reduce technology	K2

E. Correlation of COs with POs:

Cos	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	**M**	**M**	**L**	**M**		**M**						**L**	**M**		**L**
CO2	**M**	**M**	**L**	**M**									**M**		**L**
CO3	**M**	**L**	**L**	**L**								**L**	**M**		**M**
CO4	**M**	**L**	**M**	**L**								**M**	**M**	**L**	**M**
CO5	**L**	**M**	**M**	**M**	**M**							**L**	**L**	**M**	**M**

H- High; M-Medium; L-Low

F. Course Content

UNIT I Big Data Analytics **9**

Big Data Overview - State of the Practice in Analytics - Data Analytics Lifecycle - Phase 1: Discovery - Phase 2: Data Preparation - Phase 3: Model Planning - Phase 4: Model Building - Phase 5: Communicate Results - Phase 6: Operationalize

UNIT II Data Analytic Methods Using R **9**

Introduction to R - Exploratory Data Analysis - Statistical Methods for Evaluation - Advanced Analytical Theory and Methods: Clustering

UNIT III Association Rules & Regression **9**

Association Rules - A priori Algorithm - Evaluation of Candidate Rules - Applications of Association Rules - Validation and Testing. Regression - Linear Regression – Use Cases - Model Description - Logistic Regression - Use Cases - Model Description

UNIT IV Classification & Time Series Analysis **9**

Decision Trees - Naive Bayes - Diagnostics of Classifiers- Additional Classification Methods. Overview of Time Series Analysis - ARIMA Model- Additional Methods

UNIT V Text Analysis & Technology **9**

Text Analysis Steps - Collecting Raw Text - Representing Text – TFIDF. Analytics for Unstructured Data: MapReduce, Apache Hadoop

Total: 45 Hours

G. Learning Resources

Text Book

1. Data Science & Big Data Analytics Discovering, Analyzing, Visualizing and Presenting Data EMC Education Services.